DATA ANALYST INTERNSHIP

Task 5: Exploratory Data Analysis (EDA)

Observations for Each Visual

1. Pairplot (Age, Fare, Pclass vs Survived):

- Most survivors were from **1st class**.
- Passengers who paid **higher fares** had better survival chances.
- **Children (under 10)** were more likely to survive.

2. Heatmap (Correlation Matrix):

- **Fare** and **Pclass** are **negatively correlated**.
- **Survived** is positively correlated with **Fare**, and negatively with **Pclass**.
- Weak correlation between Age and Survived.

3. Histogram (Age Distribution):

- Majority of passengers were between **20 and 40 years**.
- Very few passengers were aged under 10 or over 60.

4. Histogram (Fare Distribution):

- Distribution is **right-skewed** (long tail).
- Most fares are under **$100**, with a few high outliers above **$500**.

5. Boxplot (Age vs Survived):

- Survivors tend to be **slightly younger** on average.
- Wider age range observed among survivors.

6. Countplot (Pclass vs Survived):

- **3rd class** passengers had the **lowest** survival rate.
- **1st class** passengers had the **highest** survival rate.
- Pclass is a **strong predictor** of survival.

Summary of Findings

- **Survival rates** were higher for **females, younger individuals**, and **1st class passengers**.
- **Fare amount** is an indicator of survival—higher-paying passengers had better chances.
- **Passenger class (Pclass)** is a major factor in determining survival likelihood.
- Children had better survival odds, indicating some **rescue prioritization**.
- Strong evidence of **social/economic bias** in survival patterns.